

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

EXHIBIT *17* 18

EWALD BECK AND BARBARA ZINK

GENE 16: 35-58, 1981

THIS PAGE BLANK (USPTO)

Nucleotide sequence and genome organisation of filamentous bacteriophages f1 and fd

Restriction maps; DNA sequences; genes; reading frames; regulatory signals; near identity of f1 and M13 phages)

Wald Beck and Barbara Zink

Mikrobiologie, Universität Heidelberg, Im Neuenheimer Feld 230, 6900 Heidelberg (F.R.G.)

Received April 6th, 1981)
Accepted August 4th, 1981)

SUMMARY

The DNA sequence of the filamentous phage f1, consisting of 6407 nucleotides, has been determined. When compared with the DNA sequence of the related filamentous phage fd (Beck et al., 1978), the f1 sequence is 180 nucleotides shorter and differs in 180 positions from the fd DNA. Only ten of these base exchanges cause amino acid exchanges in the known gene products. Most of the exchanges in f1 are the same as in M13 (Van Zeebeek et al., 1980), showing a near identity of these two phages (there are only 59 nucleotide differences). Regulatory units for replication, transcription, and translation are in their essential parts identical in all three phages.

INTRODUCTION

The genomes of the filamentous *Escherichia coli* phages, e.g. fd, f1, and M13, consist of single-stranded circular DNAs of about 6400 nucleotides. These code for at least nine genes, whose products are involved in phage DNA replication, phage assembly and phage protein synthesis (Marvin and Hohn, 1969; Ray, 1977). Since the propagation of the filamentous phage is catalysed mainly by host functions, their genomes have been studied for many years as model systems of regulation in *E. coli* for replication,

transcription, and translation (recent reviews Ray, 1977; Schaller, 1979). In addition, the DNA of these phages was also used early as model system in the development of methods for the structural analysis of genomes. Ling (1972) sequenced a number of large pyrimidine oligonucleotides; Oertel and Schaller (1972) determined the sequence and the order of pyrimidine tracts in a pyrimidine rich segment of the fd DNA, and Sanger et al. (1973; 1974) deduced the sequence of 89 nucleotides in the f1 DNA (pos. 6321-6408) using the ribo-substitution technique. Three ribosome-binding sites were sequenced by Pieczek et al. (1974) and the promoter site of gene X by Schaller et al. (1975) and Sugimoto et al. (1975). The DNA from the origin of replication, first isolated and characterized from a pre-initiation complex (Schaller et al., 1976) was the first continuous stretch of fd DNA to be analysed (Gray et al.,

Abbreviations: bp, base pairs; IG, intergenic region; pos., position; pDNA, DNA protected by RNA polymerase against nucleolytic DNase digestion; RF, replicative form.

1978), using the rapid methods for DNA sequencing newly developed by Sanger and Coulson (1975) and Maxam and Gilbert (1977). The DNA sequence of the origin of replication of defective interfering particles of f1 was also analysed (Ravetch et al., 1979). Takanami et al. (1976) and Sugimoto et al. (1977) analysed the region of the genes VII and VIII and the central terminator of transcription, sequencing RNA produced by in vitro transcription of restriction fragments. In 1978 the total fd DNA sequence was determined and published first in a preliminary version (Schaller et al., 1978), followed by the final sequence in a short publication (Beck et al., 1978). At that time about 90% of the DNA sequence of the related phage f1 also had been analysed mainly to confirm the gene reading frames by identifying the numerous silent base exchanges between the f1 and fd DNA. In this paper we discuss the experimental details of the fd and f1 DNA sequence analysis and the derived structures of genes and regulatory signals. In addition, we present the completed f1 DNA sequence. It differs by 180 base exchanges from the fd DNA sequence, only few of which cause amino acid changes in the gene products. As the DNA of the other closely related filamentous phage M13 has also been sequenced completely (Van Wezenbeek et al., 1980), a comparison of the three sequences is presented.

MATERIALS AND METHODS

(a) Bacteriophage and enzymes

The wild-type bacteriophage f1 and the f1 nonsense mutants amR5, amR7, amR124, and amR143 were from N.D. Zinder, New York. The bacteriophage fd was from H. Hoffmann-Berling, Heidelberg. The fd strain 478 which was isolated as a single plaque from the fd stock and used in the sequence analysis differs in at least one position (1859) from the fd phage from ATCC which was sequenced in part in the laboratory of M. Takanami, Kyoto. The viral DNA was converted into the double-stranded form (RF) in vitro by oligonucleotide primed synthesis as described (Gray et al., 1978). The restriction endonucleases *HpaII*, *HaeIII*, *HinfI*, *HhaI*, *HgaI*, *AluI*, and *TaqI* were prepared essentially as described by Roberts et al. (1976). *AccII*, *HphI*, and

MboII were purchased from New England Biolabs, and *Sau3A* was a gift from H. Streeck, Munich. Polynucleotide kinase and calf intestinal phosphatase were from Boehringer GmbH, Mannheim. [γ - 32 P]-ATP (spec. act. approx. 6000 Ci/mmol) was prepared as described by Johnson and Walseth (1979).

(b) 5'-End-labeling of DNA

Restriction fragments were dephosphorylated either by adding phosphatase into the cleavage mixture together with the restriction endonuclease, or in cases of flush-ended or 3'-extended ends in 50 mM Tris pH 8 at 60°C (0.02 units phosphatase per 20 μ l assay; incubation time 30–60 min). The samples were phenol-extracted, desalted on a small Sephadex G75 column (2 ml disposable pipette) in 10 mM ammonium-bicarbonate pH 8.6 and lyophilised. This was found to be the best method for complete removal of the phosphatase. Phosphorylation with [γ - 32 P]ATP and polynucleotide kinase was carried out essentially as described by Maxam and Gilbert (1980). In general 1–2 pmol cleaved RF DNA were used per assay.

(c) DNA sequencing methods

Gel electrophoresis, elution of DNA from polyacrylamide gels, separation of labeled fragment ends either by a secondary restriction enzyme cleavage or by separation of denatured strands, and the base-specific chemical modification were performed essentially as described by Maxam and Gilbert (1980). The depurination was carried out in 66% formic acid for 2–8 min at 20°C, followed by 3-fold dilution with water, three ether extractions, lyophilisation and hydrolysis in 1 M piperidine at 90°C for 1 h in an oven. Some fragments analysed on long (1 m) sequencing gels (0.4 mm thick) could be read up to position 450.

RESULTS AND DISCUSSION

(a) Sequencing strategy

For the complete analysis of the DNAs of fd and f1 ten different restriction endonucleases were used

(see Fig. 1). Usually the DNA was cleaved with a particular restriction enzyme and the resulting fragments were end-labeled as a mixture and separated on polyacrylamide gels. Most of the radioactive fragments were used for the sequence analysis. Many restriction maps (e.g. from *HhaI*, *HinfI*, *HphI*, *MboII*, *Sau3A*, and *TaqI*) were not established prior to sequencing but resulted from matching overlapping sequences.

In the case of restriction endonucleases *HgaI* and *TaqI*, which each have only ten cleavage sites in the fd DNA, the restricted and end-labeled DNA was further cleaved by a second restriction enzyme before separation on a gel. By comparing fragments present before and after the second cleavage it could be deduced which new fragments had been generated by the second digestion and which were thus labeled at only one end and which could be used directly in the subsequent analysis. Using this method separation of the re-cleaved fragments on a second gel was unnecessary. However, there was often a higher degree of contamination by neighbouring bands or background which interfered with extended reading of nucleotide sequences.

The restriction enzyme *HaeIII* also cleaves single-stranded DNA efficiently (Blakesley and Wells, 1975). With this enzyme single-stranded fragments could be prepared directly and used for sequencing without secondary cleavage.

Although it was usually possible to read the DNA sequences clearly, 85% of the fd DNA was sequenced in both strands to avoid mistakes that could occur at methylated bases (Ohnori et al., 1978), in regions with a distinct secondary structure, or by incorrect reading or processing of the sequence information. Care was taken that all restriction sites used to generate fragments were read through from alternative starts. This is particularly important in repetitive sequences that may contain closely spaced repeating restriction sites. Such an example occurs in the fd DNA sequence around position 2390, where a sequence of 18 nucleotides consisting of two small *HpaII* fragments was not included in the preliminary version of the fd DNA sequence (Schaller et al., 1978). The nucleotide sequences were stored and processed, using computer programmes written in the computer language APL and established by Osterburg and Sommer (1981).

(b) The DNA sequence

The fd sequence was derived by reading serially overlapping fragments, making use of most of the different restriction cuts and both DNA strands as shown in Fig. 1. Sequencing of f1 DNA was started somewhat later. Therefore, only about 50% of the sequence analysis was carried out in both strands, since we could refer to the completed fd sequence. Nevertheless, some regions of f1 were analysed in more detail than in fd, and the f1 sequence was also used to confirm fd DNA sequences which had been determined in one strand only.

Fig. 2 shows combined sequences of the fd and the f1 DNA. The continuous sequence corresponds to the fd DNA sequence as published in 1978 (Beck et al.). About 97% of the f1 DNA is identical to the fd DNA. There exist 180 base changes, which are indicated above the fd DNA sequence. Whereas about 150 of them lie within genes, only 10 actually cause amino acid changes. The others are "silent" alterations, i.e., they involve variable bases in the codons. This fact was used already earlier as indirect evidence for the correct reading frames of the genes in the filamentous phage genome (Schaller et al., 1978; and see below). Base changes present in the M13 DNA sequence (Van Wezenbeek et al., 1980) are also included in Fig. 2. Many of them coincide with the changes in f1, demonstrating that f1 and M13 are more closely related to each other than to fd.

A series of partial sequences from fd DNA and f1 DNA published earlier (see above) could be fitted into the complete sequences. All agree essentially with our data. Two changes had to be made in regulatory regions: one at the promoter of gene VIII, where the sequence at the start of transcription is a G₃ run (not G₄ as in Takanami et al., 1976), the other at the central terminator, where the sequence at the end point of transcription is C₂T₉ (not C₂T₈C as in Ling, 1972, or C₂T₈ as in Sugimoto et al., 1977). An f1 DNA sequence of the intergenic region (IG) between genes IV and II (position 5500–6000; Fig. 3) analysed by Ravetch et al. (1977; 1979) contains several deletions of one or two nucleotides when compared with the corresponding fd DNA sequence. None of these deletions could be confirmed in our f1 DNA sequence. The corresponding region in M13, analysed first by Suggs and Ray (1978) and confirmed by the M13 DNA sequence of Van Wezenbeek

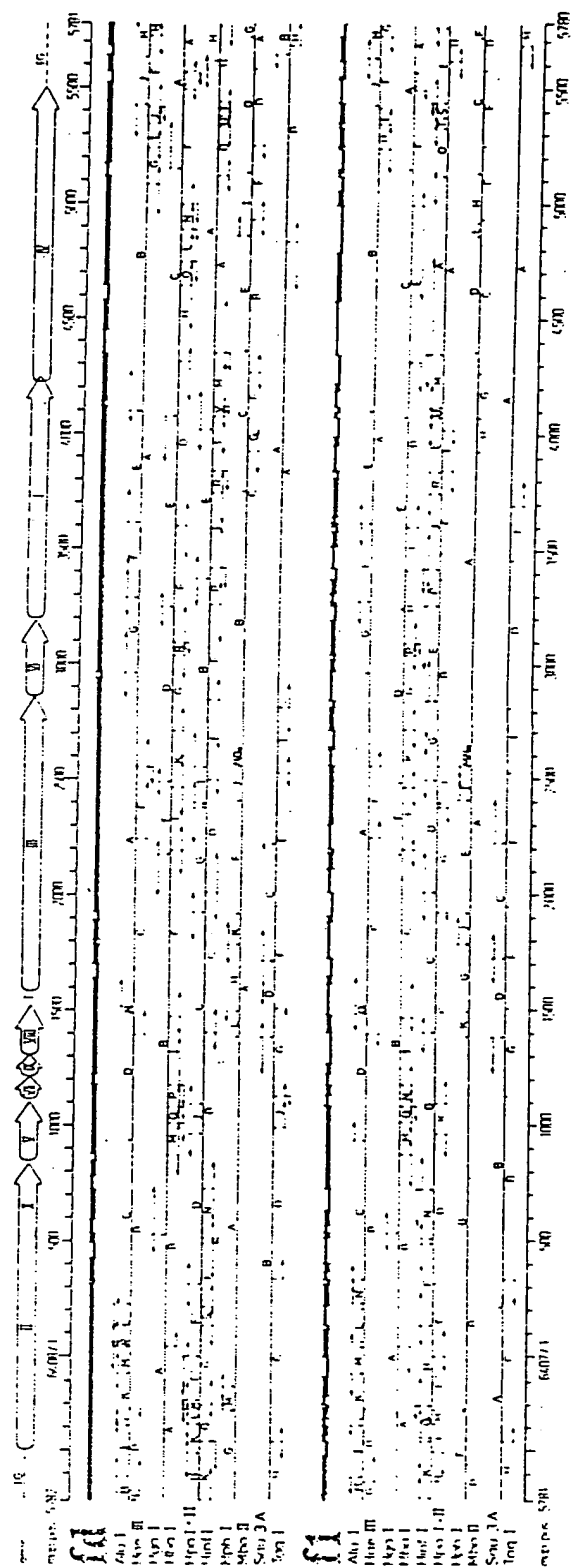


Fig. 1. Genetic and physical linearized maps of bacteriophages fd and f1, including the sequencing strategy applied. The circular phage genome is opened at the start point of viral strand replication in the intergenic region between genes IV and II (IG). The genes (roman numerals) and the central terminator of transcription (T) are indicated in the top line. The zero point of the map (see 6408/1 and 6407/1 plus positions) is the single *Hind*III (*Hpa*I) cleavage site in fd. The bars for fd and f1 each indicate how far the DNA sequence is determined in both strands (solid bar) or in the minus strand only (gap in the lower part of the bar) or in the minus strand only (gap in the upper part of the bar). The length and the orientation of the individual sequencing runs are indicated by arrows in the restriction enzyme cleavage maps.

al. (1980) agrees with the f1 sequence except for 3 positions. Whereas M13 and f1 are almost identical in this region, fd differs in 23 positions of the antigenic region from these two phage.

In addition to the published results, other fd DNA sequence data for the region between position 300-1600 were made available (M. Takanami, personal communication). In this analysis a difference between our fd DNA and that used by Takanami was noticed: a G → A exchange at position 1859 creating additional *HinfI* site in Takanami's DNA. The red restriction fragment pattern of this enzyme demonstrated experimentally (M. Takanami, personal communication).

Restriction maps

When the work on fd sequence analysis started in 1977, several restriction maps (*HpaII*, *HaeIII*, *AluI*, *EcoRI*) had already been completed for fd, f1 and/or f2. During the sequence analysis these maps were revised and maps for many other restriction enzymes published. The maps of *TaqI*, *HhaI*, *HinfI*, *HaeIII*, *BamHI*, *HphI*, *MboII*, and *AccII* (*ThaI*) were checked experimentally by comparison of the fragment length derived from the DNA sequence with the corresponding fragment patterns on polyacrylamide gels. The recognition sites in the three filamentous phage DNAs of the best known restriction enzymes listed in Table I. In nondenaturing polyacrylamide gels some fragments (e.g. in fd *HpaII*-B (pos. 2-3371), *HpaII*-H (pos. 5615-5996) and *TaqI*-H (pos. 5648-6041) migrate more slowly than other fragments of comparable length. Such fragments usually contain extended inverted repeats, which may be secondary structures divergent from the normal double helical form of DNA.

Genes and gene products

A genetic map of the eight known genes of the filamentous phage was established by Lyons and Der (1972) and correlated later with the size of gene products determined on SDS gels (Model and Zinder, 1974) and the physical maps (Voivis et al. 1975). In addition to these approximate positions of gene lengths the amino acid sequences of gene V and gene VIII proteins (Nakashima and Koningsberg, 1974; Nakashima et al., 1974) were determined,

TABLE I

Restriction endonuclease recognition sites

The sites for fd, f1 and M13, as found by computer analysis, are compiled (c.f. also Fuchs et al., 1980). Italic numbers represent cleavage sites that are experimentally proven. No cleavage sites exist in all three phages for the enzymes *AvaII*, *AvaIII*, *BclI*, *BglI*, *BglII*, *EcoRI*, *HindIII*, *KpnI*, *MsrI*, *PstI*, *PvuII*, *SacI*, *SacII*, *Sall*, *SmaI*, *XhoI*, and *XbaI*.

| Name | Sequence | Position |
|---------------|----------------------------------|---|
| <i>ACEI</i> | GTACAC | 609 ^a |
| <i>AluI</i> | AGCT | 38 83 203 ^{bc} 229 333 ^{bc} 336 1498 1517 2262 2277 3613 4057 5427 5621 5888 6101 6131 6336 |
| <i>AsuI</i> | GGCC | 5725 |
| <i>AvaI</i> | CTCGGG | 5826 ^c |
| <i>BalI</i> | TGGCCA | 5081 ^{ac} |
| <i>BamHI</i> | GGATCC | 2220 5615 ^b |
| <i>BbvI</i> | CC ^a CC | 932 1367 2521 3132 4872 5537 |
| <i>ClaI</i> | ATCGAT | 2527 ^{ac} 6040 |
| <i>DdeI</i> | CTAG | 233 1099 1371 1417 1784 1847 1862 1879 1901 1973 2015 2318 2333 2348 2363 2676 ^a 3362 4014 4041 4080 4094 4122 4282 ^c 4882 ^a 5263 5371 ^b 6066 ^{bc} 6219 ^{bc} 6347 |
| <i>EcoB</i> | TGATN ₁ TCGT | 559 ^{ab} 2348 |
| <i>EcoRII</i> | CC ^a GG | 1014 1966 |
| <i>EcoRV</i> | GC ^a GC | 932 1367 1394 1871 ^b 2285 2288 2312 2327 ^a 2357 2384 ^{ab} 2521 3132 4872 4888 ^a 5501 5515 5537 |
| <i>HaeII</i> | AGCGCT | 2710 ^{ac} 3036 ^{bc} 4743 ^a 5897 5997 |
| <i>HaeIII</i> | GGCC | 1395 2245 2254 5097 ^{ac} 5247 5249 5672 5732 5823 ^a 5867 ^{bc} 6187 |
| <i>HhaI</i> | CACCG, CGCTC | 327 3164 2473 2121 4274 4253 |
| <i>HhaII</i> | C ^a CC ^a C | 4744 ^{bc} 5466 |
| <i>HhaI</i> | GGCC | 44 812 ^{ab} 1021 1285 1377 1394 2497 4701 5760 5896 5923 ^{bc} 5939 6223 6247 6744 ^a 6832 ^a 6895 6957 5274 5277 5823 5997 6078 |
| <i>HindII</i> | GTTAC | 4713 ^b 6024 |
| <i>HinfI</i> | CAATC | 125 225 ^{ac} 487 511 573 1470 ^{ab} 1522 2467 2664 2689 3473 3741 3829 4272 4274 4287 5181 5220 5375 5423 5977 5978 6042 6062 ^{bc} 6186 6428 |
| <i>HpaI</i> | GTTAAC | 6420 |
| <i>HpaII</i> | CCGG | 224 925 1291 1824 2373 2382 ^{bc} 2587 2811 3372 3443 ^{bc} 4623 4625 4991 5119 5179 6282 ^a |
| <i>KpnI</i> | GGTGA, TCACC | 1276 1362 1774 1923 2299 3142 3391 3822 2520 2635 3740 ^a 4247 ^{ab} 4251 4945 4946 ^{bc} 5172 5727 6182 6239 6250 ^a |
| <i>MboII</i> | GAAGA, TCTTC | 787 ^{bc} 2128 ^b 2822 4275 4277 4935 5239 5387 5967 ^{bc} |
| <i>MnlI</i> | CCTC, GAGC | 254 331 ^a 373 484 560 587 625 655 782 ^a 1031 1088 1231 1297 1318 1326 1345 1373 1416 1506 1663 ^a 1732 1834 1849 1864 1879 1897 1945 2008 2020 2218 2263 2269 2320 2325 2350 2365 2369 ^c 2673 ^c 2677 2894 3052 3222 3337 3353 3704 4022 4308 ^{ac} 4399 ^a 4699 4773 ^{bc} 4821 ^{bc} 4836 4922 ^{bc} 4927 5037 ^a 5248 5417 5448 5682 5688 ^a 6096 6114 6244 6349 |
| <i>SmaI</i> | GTAC | 173 280 1022 1165 1769 1794 1889 1905 1976 2133 3468 3669 ^{bc} 4191 4381 5385 ^{bc} 5462 ^a 5487 6001 6323 6390 ^a |
| <i>SmaII</i> | GATC | 719 ^b 1382 1714 2727 3649 ^a |
| <i>SmaIII</i> | GATGC, GCATC | 25 388 1031 ^a 1354 3980 4851 |
| <i>ThaI</i> | CCGG | 43 347 1119 1176 2466 2710 ^b 3356 3410 ^{bc} 3600 3953 4314 4425 ^a 4641 ^a 4887 ^a 4993 5490 5514 5534 5910 |
| <i>TaqI</i> | TCCA | 327 399 ^a 1117 1522 1943 2127 2622 ^{ab} 2630 ^{bc} 3691 ^{bc} 4664 ^a 4774 ^a 5789 5947 |
| <i>XhoII</i> | CGATCT | 215 ^b 2270 5445 ^a |

^a Site exists in fd only; ^b site exists in f1 only; ^c site exists in M13 only.

[illegible]

Arg61GlyLysProTyrSerLeuAsnGluGlnLeuCysTyrValAspLeuGlyAsnGluTyrProValLeuValLysIleThrLeuAspGluGlyGln
C T M C

1001 GTCACGGCATCCGCCCTGGCTGTATACACCCTGCATCTGCTTGCTTCAGAGTTGGTTCAGTTCGGTCTCTTAATGATGACCGCTCGCGCCCTGATTCGGACT

R1 R2 R3

ProValTyrAlaProGlyLeuTyrThrValHisLeuSerSerPheValLysValGlyGlnPheGlySerLeuMetIleAspArgLeuArgLeuValProAla
gene V end gene VII start

1101 AGCAAAATGAGTGGCACGGTTCGGATTTCACACAAATTATACAGCTGATGATGACAATCTCCGTGTACTTGTTCGGCGCTGGTATAATCGCTGGGGGT
MetGlnGlnValAlaAspPheAspThrIleTyrGlnAlaIleGlnIleSerValLecucyPheAlaLeuGlyIleIleValGlyGly

R1 R2 R3

Lys*** gene IX start
gene VII end

1201 CAAGGATGAGTGGTTTAGTGTATCTTCGGCTCTTCGTTTAGGTGGTGGCTTCGTAGTGGCATTAAGATTTACCCGTTAATGGAACCTTCCTC
GlnArg***MetPhe***
R1 R2 R3

MetSerValLeuValTyrSerPheAlaSerPheValLeuGlyTyrCysLeuArgSerGlyIleThrTyrPheThrArgLeuMetGluThrSerSer
gene VIII start I

1301 ATGAAAAGTCTTAGTCTCAAGGCTCCGTACCGCTGCTACCCCTGCTCCGATCTGCTTCGGTGGAGGGTGAAGCATCCGCGCAAAGCGGCT
MetLysLysSerLeuValLeuLysAlaSerValAlaValAlaThrLeuValProMetIleuSerPheAlaAlaGluGlyAspProAlaLysAlaAlaPhe
M ***
R1 R2 R3

TTCATCTCCCTGCAAGGCTCGACGACCGCATATATCGGTATGCTGGGCTATGGTGTGTGCATTTGCGGCCAACATATCGGATCAAGCTGTTAAGA
ASP SerLeuGlnAlaSerValThrGluTyrIleGlyTyrAlaTrpAlaMetValValValIleValGlyValAlaThrIleGlyIleLysLeuPheLysLys

R1 R2 R3

gene VIII end
1501 ATTACCTCGAAAGCAGGGTGTATGAGCCCATACAAATTAGAGGCTCTTTTGGAGCTTTTTTTTGGAGATTTCACGCTGAAAAAATATTATTCGCA
PheThrSerLysAlaSer*****
MetLysLysLeuLeuPheAlaIle

R1 R2 R3

gene III start
1601 TTCTTTAGTGTCTCTTTCTTCTTCTCACTCCGCTGGAAAGTGTGAAAGTGTGTAGCGAAACCTCATACAGAAATTTCAATTACTAACTCTGGAAAGA
ProLeuValValProPheTyrSerHisSerAlaIleThrValGluSerCysLeuAlaLysProHisThrGluAsnSerPheThrAsnValTrpLysAsp

R1 R2 R3

1701 CGACAAACATTGATCGTTACGCTAACCTATAGAGGCTGCTCTGTGGAACTCTACAGGCGTGTGGTTGTACTGGTGAAGAACTCAGTGTACCGGTACA
AspLysThrLeuAspArgTyrAlaAsnTyrGluGlyCysLeuTyrPheAlaThrGlyValValLysThrGlyAspGluThrGlnCystyrGlyThr

R1 R2 R3

1801 TGGGTCTTATGGCTTGTATCTCCCTGGAATGAGGATGGTGGCTCTGAGGGTGGCGGTCTGAGGGTGGCGGTCTGAGGGTGGCGGTACTAAACCTC
TrpValProIleGlyLeuAlaIleProGluAsnGluGlyGlyGlySerGluGlyGlySerGluGlyGlySerGluGlyGlyGlyThrLysPro

R1 R2 R3

[illegible]

 PheLeuLeuValThrLeuPheGlyTyrLeuLeuThrPheLeuLysGlyPheGlyLysIleAlaIleSerLeuPheLeuAlaLeuIleGly
 3001 GCGTTAACTCAATTCCTTGGGTTATCCTCCTGATATTACCGCACATACCTGCTGATTTTGGTTCAGGGCTTACGTTAACTTCCGGCTAAATGCGCT

 LeuAsnSerIleLeuValIGlyTyrLeuSerAspIleSerAlaGlnLeuProSerAspPheValIGlyValGlnLeuIleLeuProSerAsnAlaLeu
 3101 TCCCTGTTTTATGTTATTCCTCTCTGTAAGGCTGCTATTTTCATTTTTCAGCTTAACAGAAATCGTTCTTATTTGGATTGGGATTAATAATATGG

 ProCysPheTyrValIleLeuSerValLysAlaAlaIlePheIlePheAspValLysGlnLysIleValSerTyrLeuAspTrpAspLys*** MetAla
 3201 CTGTTTATTTTGTAACTGGCAAAATTAGGCTCTGGAAGAGACGCTTCTAGGCTTGGTAAGATTACGATAGAAATTTAGCTGGGTGCAAAATAGCAACTAA

 ValTyrPheValIleThrGlyLysLeuGlySerGlyLysThrLeuValSerValIGlyLysIleGlnAspLysIleValAlaGlyCysLysIleAlaThrAsn
 3301 TCTTGATTAAAGGCTTCAAAAGCTCCCGCAAGTCGGGAGGTTCCGCTAAAGCGCTCGGTTCTTAGAATACCGGATAGGCTTCTTATTTCTGATTGCTT

 LeuAspLeuArgLeuGlnAsnLeuProGlnValIGlyArgPheAlaLysThrProArgValIleuArgIleProAspLysProSerIleSerAspLeuLeu
 3401 GCTATTGGTCTGGTAAATGATTCTTACGACGAAATAGAAACGGTTTGGTTCTTGGTAAGTGGGTTACTTGGTTTAAATACCGGTTTACGGAATGACA

 AlaIleGlyArgGlyAsnAspSerTyrAspGluAsnLysAsnGlyLeuLeuValLeuAspGluCysGlyThrTrpPheAsnThrArgSerTrpAsnAspLys
 3501 AGGAAAGACAGCGGATTATGATTGGTTCTTCTTCTGATGCTCGTAAATTTGGGATGGGATATTATTTTCTTGTTCAGGATTTAICTATTGTTGATAAACAGGC

 GluArgGlnProIleLeuAspTrpPheLeuHisAlaLeuArgLysLeuGlyTrpAspIleIlePheLeuValIGlnAspLeuSerIleValAspLysGlnAla
 3601 CCGTCTGCTAGCTGACACAGCTTGTATTATTGTCGCGCTCGACAGAAATTACTTTACGCTTGGTGGCCACTTATATCTCTGTTACTGGCTCAAAA

 ArgSerAlaLeuAlaGlnHisValValTyrCysArgArgLeuAspArgIleThrLeuProPheValIGlyThrLeuTyrSerLeuValThrGlySerLys
 3701 ATGCTCTGCTAATACATGTTGGTGTGTTGTTAATATGTTGATTCATGATTAAGGCTTACGCTTGGCTTATGTAAGAAATTTATATA

 MetProLeuProLysLeuHisValIGlyValValLysTyrGlyAspSerGlnLeuSerProThrValIGluArgTrpLeuTyrThrGlyLysAsnLeuTyrAsn
 3801 ACCGATATGACACTAAGAGGCTTTTTCAGTAAATATGATTCAGGTTGTTATTCATATTAACCCCTTATTATCAGACAGGCTGGTATTTCAACCCATT

R1
 R2
 R3

 AlaTyrAspThrLysGlnAlaPheSerSerAsnTyrAspSerGlyValTyrSerTyrLeuThrProTyrLeuSerHisGlyArgTyrPheLysProLeu

 3901 AAATTTAGGTCAGAGATGAAATTAACATAAAATATATTGGAAAGAGTTTTCGCGGTTCTTTGGCTTGGGATAGGATTTGGCATCAGCATTACATATAGT

 R1
 R2
 R3
 AsnLeuGlyGlnLysMetLysLeuThrLysIleTyrLeuLysPheSerArgValLeuCysLeuAlaIleGlyPheAlaSerAlaPheThrTyrSer

 4001 TATATAACCCAACTAAGCCGGAGGTTAAAGGTTAGTCTCTCAGACCTATGATTTTGGATAAATTCACCTATTTGACTTCTCTCAGCGCTTAATCTAAGCT

 R1
 R2
 R3
 TyrIleThrGlnProLysProGlnValLysLysValValSerGlnThrTyrAspPheAspLysPheThrIleAspSerSerGlnArgLeuAsnLeuSerTyr

 4101 ATCGCTATGTTTTCAGAGGATTCAGGGAAATTAATTAATAGCGACGATTTACAGAGGCAAGGTTATTCATCACAATATATTGATTATGACTGTTTC

 R1
 R2
 R3
 ArgTyrValPheLysAspSerLysGlyLysLeuIleAsnSerAspAspLeuGlnLysGlnGlyTyrSerIleThrTyrIleAspLeuCysThrValSer

 4201 AATTAAAAAGGTAATCAAAATGAAATTTGTTAAATGTTAAATTAATTTTGTTCCTTGATGTTTGTTCATCATCTCTTTTGGCTCAAGTAATTAATGAA

 R1
 R2
 R3
 IleLysLysGlyAsnSerAsnGlnIleValLysCysAsnMetLysLeuLeuAsnValIleAsnPheValPheLeuMetPheValSerSerSerPheAlaGlnValIleGluMetAsn

 4301 TAATTCGCTCTGCGCGATTTTCGTGACTTTGGTATTCAAGCAACACAGGTAATCTGTTTATGTTCTCACCCTGATGTTAAAGGTACAGTGTATATATCC

 R1
 R2
 R3
 AsnSerProLeuArgAspPheValThrTyrTyrSerLysGlnThrGlyGluSerValIleValSerProAspValLysGlyThrValThrValTyrSer

 4401 TCTGACGTTAAGCTGAAATTTACGCAATTTCTTATCTCTGTTTACGCTGCTAATAATTTTGTATATGTTGGCTCAATTCCTCCATAATTCAGAAAT

 R1
 R2
 R3
 SerAspValLysProGlnAsnLeuArgAsnPhePheIleSerValLeuArgAlaAsnAsnPheAspMetValGlySerIleProSerIleIleGlnLysTyr

 4501 ATACCCCAATATGTCAGGATTAATTAATGAAATTTGCCATCATCTGATATTCAGGAAATATGATGATAATTCGCTCTCTGTTGGTGTTCCTTTGTTCCGCA

 R1
 R2
 R3
 AsnProAsnSerGlnAspTyrIleAspGluLeuProSerSerAspIleGlnGluTyrAspAspAsnSerAlaProSerGlyGlyPhePheValProGln

 4601 AATGATTAATGTTACICAAACATTTTAAATTAACGTTTCGCGCAAGGATTTAATAGGGTTGTAGGATTTGTTTAAATCTAAATCTAAATCC

 R1
 R2
 R3
 AsnAspAsnValThrGlnThrPheLysIleAsnAsnValArgAlaLysAspLeuIleArgValValGlnLeuPheValLysSerAsnThrSerLysSer

 4701 TCAGATGATATATCTGTTGATGGTTCTTAACCTATTATAGTAGT, GCGGCCCTTAAGATATTTAGATAACTCCGCAATTTCTTCTACTGTTGATTGCT

SerAsnValLeuSerValAspGlySerAsnLeuLeuValValSerAlaProLysAspIleLeuAspAsnLeuProGlnPheLeuSerThrValAsnLeuPro
 4801 CAACGACCAAGATATTGATTGAAGGATTAAATTTTCGAGGTCAGCAAGGTGATGCTTTAGATTTTCCCTTGGCTCTCAGCGCGGCACTGTTGC

 ThrAspGlnIleLeuIleGluGlyLeuIlePheGluValGlnGlnGlyAspAlaLeuAspPheSerPheAlaAlaGlySerGlnArgGlyThrValAla
 4901 TGGTGGTGTAAATACAGCCGCTCAACCTCIGTTTTATCTCTCGCGGGTGGTTCCGTTCCGTTATTTTACGGCGATGTTTTAGGGCTATCAGTTCCGCGCA

 GlyGlyValAsnThrAspArgLeuThrSerValLeuSerSerAlaGlyGlySerPheGlyIlePheAsnGlyAspValLeuGlyLeuSerValArgAla
 5001 TTAAAGACCTAGTACCCATTCAAAAATATTGCTGTGCTGCTGATTTCTTACGCTCAGAGGGTCTATTTCTGTTGGCCAGAAATGTCCTTTTA

 LeuLysThrAsnSerHisSerLysIleLeuSerValProArgIleLeuThrLeuSerGlyGlnLysGlySerIleSerValGlyGlnAsnValProPheIle
 5101 TTAATCGTCTGTTAACTGCTGAATCTGCCAATGTAAATAATCCATTTTCAGCGGTGAGCGTCAAAATGTTGGTATTTCATGAGTGTTCCTCCGTTGC

 ThrGlyArgValThrGlyGluSerAlaAsnValAsnProPheGlnThrValGluArgGlnAsnValGlyIleSerMetSerValPheProValAla
 5201 AATGGCTGGCGGTAAATATTGTTTTAGATATAACCGATAGCGCGGATAGTTTGAGTCTCTTCTACTCAGGCAAGTGAATGTTATTACTAATCAAGAGATATT

 MetAlaGlyGlyAsnIleValLeuAspIleThrSerLysAlaAspSerLeuSerSerThrGlnAlaSerAspValIleThrAsnGlnArgSerIle
 5301 GCGACACCGGTAAATTTGGGTGATGGTCAAGACTCTTTTGGTGGTGGCTCAGTATTACAAAAACACTTCTCAAGATTCGTGGTGGCCGTTCTGCTGCTA

 AlaThrThrValAsnLeuArgAspGlyGlnThrLeuLeuLeuGlyGlyLeuThrAspTyrLysAsnThrSerGlnAspSerGlyValProPheLeuSerLys
 5401 AAATCCCTTTAATCGGCCCTCTGTTTACGCTCCGTTCTGATTCACGAGGAAAGCAGCTGTGACGTGCTGCTCAAGCAACCATAGTACGCGGCTGTA

 IleProLeuIleGlyLeuLeuPheSerSerArgSerAspSerAsnGluGluSerThrLeuTyrValLeuValLysAlaThrIleValArgAlaLeu***
 5501 GCGGCGCATTAAGCGCGCGGGTGTGGTGGTTACGCGCGAGCTGACCTACACTGGCCAGCGCCCTAGCGCCCTCCTTTCCGTTCTCTCCCTTCCTT

 G
 5601 TCTCGCCACGCTCTCCGGCTTCCCGGTCAGGCTCTAAATCGGGGGATCCCTTTAGGGTTCGGATTTTAGGGCACCTCGACCTCCAAAAACTT
 C A

*
*
*

Fig. 2. Nucleotide sequence of the DNA of bacteriophage fd. Base exchanges for f1 and M13 are indicated; no brackets = exchange is common for f1 and M13; parentheses () = exchange exists for f1 only; brackets [] = exchange exists for M13 only. The starts and stops and the amino acid sequences for the gene products are shown. Stop codons in the three reading frames are indicated as asterisks. The lines represent possible reading frames of more than 30 amino acid residues.

allowing an exact correlation of these genes with the nucleotide sequence. The start sites of two genes were determined, or later on confirmed, respectively, by sequence analysis of the N-terminal amino acids of the proteins (gene III: Goldsmith and Koningsberg, 1977; gene II: Meyer et al., 1980). Start sites of three other (unknown) genes were determined by three ribosome-binding sites sequenced by Pieczenik et al. (1974). Two of these sites were subsequently correlated to genes V and VIII, the best characterized products of the filamentous phage. The third site defines the start of gene IV.

The exact positions of the genes are shown in the final nucleotide sequence (Fig. 2). There is only one possible reading frame for each gene within the limits derived from the genetic map. The other reading frames contain many stop codons. Continuous reading frames longer than 30 amino acids that are not allied with known genes are also indicated in Fig. 2. In most cases translation is made unlikely by the absence of a Shine-Dalgarno sequence except for two theoretical peptides of 65 and 42 amino acids starting at positions 3417 and 4528, respectively. There is no genetic evidence for additional genes in the filamentous phage.

The arrangement of the genes reflects an economical usage of DNA. The only two noncoding regions between genes VIII and III and between genes IV and II contain the central terminator of transcription and the origins of DNA replication, respectively. Gene IX overlaps with one nucleotide at each end with the neighbouring genes VII and VIII. Most other genes are separated by one or two nucleotides from each other. These one or two nucleotides obviously function to change the reading frame for the following gene and to avoid the synthesis of fusion proteins in case of readthrough by suppression. This principle is also demonstrated by the deletion of one of the two nucleotides in the intergenic space between genes VI and I in the f1 and M13 DNA compared with the fd DNA (pos. 3195). In both cases a change of the reading frames is maintained. In contrast to the eicosahedral phages (ϕ X174, G4), there are no gene overlaps of more than one nucleotide in the filamentous phage genome except of a short run of 20 bases, which is common for genes I and IV (pos. 4221-4240). There is obviously no selection pressure to limit the genome length of these phages. By insertion of heterogeneous DNA into the genome of fd,

hybrid phages could be constructed which were several times longer than wild-type fd (Herrmann et al., 1978).

Since mistakes can occur in establishing a DNA sequence for different reasons – as mentioned above – it is necessary to have useful criteria to control the derived reading frame of the genes. The best controls are data from protein sequence analysis. Amino acid sequences, partial or complete, were available for the genes V, VIII, III, and II. Another control that contributes greatly to the credibility of the complete sequence derived from the DNA, amino acid composition, was available for the gene III product.

There are also several possible ways to check the correctness of a reading frame at the nucleotide level. A *first* simple possibility is based on the fact that about 50% of all triplets within the genes end with a T residue similarly to the eicosahedral phage ϕ X174 and G4 (Sanger et al., 1977; Godson et al., 1978). The filamentous phage mostly use codons with the highest number of T's for all amino acids (see Table II). Although it cannot be used as an exact proof for the correctness of a sequence in a specific short region, this phenomenon can be used to confirm the reading frame over a longer distance. *Secondly*, the filamentous phage obviously do not have overlapping genes. There are in the unused reading frames stop codons every 30–40 nucleotides on the average. A *third* test used exists in a comparison of the DNA sequences of the closely related phages fd and f1: we determined 280 base exchanges, 120 of them within the genes. Only ten of these result in amino acid exchanges, the others are "silent" in the correct reading, i.e. they concern the variable bases in the codons. The *fourth* and most conclusive but also most elaborate method used is the determination of base exchanges to amber mutants. For almost all genes the DNA sequence of one or several amber mutants was analysed (see Table IV).

The genes are arranged in three functional groups in the genome: replication (genes II and V), capsid (genes IX, VIII, III, and VI), and morphogenesis (genes I and IV). According to this, the gene VII protein (unknown function) could either be involved in replication or be part of the virion from its position on the DNA.

The most significant functional and biochemical features of the gene products and the criteria for

TABLE II

Codon usage in fd

| | | | | | | | | | | | |
|-----|-----|----|-----|-----|----|-------|-----|----|------|-----|----|
| Phe | TTT | 67 | Ser | TCT | 92 | Tyr | TAT | 65 | Cys | TGT | 16 |
| | TTC | 39 | | TCC | 33 | | TAC | 14 | | TGC | 8 |
| Leu | TTA | 65 | | TCA | 35 | ochre | TAA | 5 | opal | TGA | 3 |
| | TTG | 32 | | TCG | 9 | amber | TAG | 1 | Trp | TGG | 18 |
| | CTT | 49 | Pro | CCT | 46 | His | CAT | 12 | Arg | CGT | 32 |
| | CTC | 17 | | CCC | 9 | | CAC | 6 | | CGC | 16 |
| | CTA | 6 | | CCA | 13 | Gln | CAA | 35 | | CGA | 5 |
| | CTG | 26 | | CCG | 18 | | CAG | 43 | | CGG | 1 |
| Ile | ATT | 72 | Thr | ACT | 60 | Asn | AAT | 82 | Ser | AGT | 14 |
| | ATC | 16 | | ACC | 23 | | AAC | 23 | | AGC | 11 |
| | ATA | 20 | | ACA | 15 | Lys | AAA | 73 | Arg | AGA | 11 |
| Met | ATG | 33 | | ACG | 11 | | AAG | 34 | | AGG | 5 |
| Val | GTT | 98 | Ala | GCT | 59 | Asp | GAT | 72 | Gly | GGT | 95 |
| | CTC | 18 | | GCC | 16 | | GAC | 38 | | GGC | 51 |
| | GTA | 25 | | GCA | 28 | Glu | GAA | 40 | | GGA | 5 |
| | GTG | 11 | | GCG | 17 | | GAG | 31 | | GGG | 9 |

localisation of the reading frames are summarized in Table III. Gene II shows two possible ATG start sites in positions 6007 and 6016. Based on a better Shine-Dalgarno sequence we have predicted that the former one must correspond to the protein start (Schaller et al., 1978). This was confirmed by determination of the N-terminus for 90% of the gene II product, using radiolabel Edman degradation (Meyer et al., 1980). However, about 10% of the protein showed amino acids in positions that correspond to a start at the second ATG codon. Whether the two proteins which were co-isolated from a membrane fraction and which are not to be distinguished on SDS gels have different biological functions is not known.

The existence of a ninth gene in the filamentous phage genome between the genes VII and VIII was predicted already from the preliminary fd sequence by Schaller et al. (1978). A gap of 94 nucleotides with no known coding or regulatory function shows a continuous reading frame, whereby the first and the last triplet each have an overlap of one nucleotide with the adjacent genes. The protein predicted from the sequence consists of 32 amino acids with a composition (6 Ser, 2 Arg, no His) that is similar to that of the C-protein, a minor capsid component which has been detected in highly purified f1 and M13 phage (Simons et al., 1979). There are no amber

mutants known for the gene IX. This is explained by the DNA sequence, which shows that possible amber codons can only be created by transversions in positions 1223, 1249, and 1274. Hydroxylaminic treatment used to construct amber mutants of the filamentous phage (Lyons and Zinder, 1972) could induce only transitions.

Gene III protein contains a remarkably high degree of glycine residues (16%). Most of these are clustered in repetitive sequences: the sequence Glu-Gly-Gly-Gly-Ser appears three times around amino acid position 95 and four times around position 255, accompanied by repetitions of Gly-Gly-Gly-Ser at both sites. In the DNA of an fd Tn5 derivative a stretch of 30 nucleotides, corresponding to two of the Glu-Gly-Gly-Gly-Ser repeats (amino acids 253-262) is deleted (Auerswald, 1979). The deleted amino acids are obviously not essential for gene III function. Around amino acid 375 the protein seems to be variable too, since base exchanges in positions 2699, 2702, and 2710 result in amino acid changes in f1 and M13 relative to fd (Table V).

M_r estimations of the gene III protein differed between 55 000 and 68 000 depending on the SDS gel system used (Goldsmit and Koningsberg, 1977). Even the lowest value differs markedly from that derived from the DNA sequence (M_r 42 660). The unusual clustering of glycine residues may alter the

binding of SDS and therefore the migration of the protein on gels.

The reading frame of gene IV extends 20 nucleotides back into the 3' end of gene I. Most of the nucleotide sequence of the ribosome binding site is homologous to the ribosome binding site of gene V. Sequence homologies are also recognizable in the first 20 nucleotides of the two genes. Moreover, parts of this homologous sequence are repeated within the coding region of gene IV (positions 3901–3931 and 4285–4305), in both cases centered around ATG codons, probably reflecting an evolutionary pathway. The reading frame of this gene ends with a TAG codon in position 5499. This stop codon lies directly at the beginning of the largest hairpin in the viral DNA. This structure may help to terminate transcription and/or translation, resulting in the correct length of the gene product even in UAG suppressor strains.

(e) Regulatory signals

Structures of regulatory signals concerning all three levels of phage development, replication, transcription and translation are recognizable on the DNA. The regulatory unit of replication lies in an intergenic region of 508 bp (= IG) between the end of gene IV and the start of gene II. This DNA segment can be folded into several large hairpin structures (Fig. 3). The existence of these structures in the viral DNA was demonstrated by their resistance to S1 nuclease (Gray et al., 1978). Their significance is indicated by the fact that their sequence is conserved in all three filamentous phage. Although the IG is the most variable region in the filamentous phage genome – more than 5% of the bases of fd differ from f1 and M13 – none of these base changes lies in the stem of a hairpin but only in the regions between. Most of the region between the end of gene IV (pos. 5501) and the hairpin C does not seem to be necessary for propagation of the DNA. By cloning parts of the IG in plasmid pBR322 under conditions not permissive for ColE1-directed replication it has been shown that positions 5727–5868, containing the start site of the *ori*-RNA (Geider et al., 1978) and the nicking site of the gene II protein (Meyer et al., 1979), is sufficient in the presence of helper virus for replication of the hybrid replicon (Cleary and Ray, 1980; Sommer, 1981). In a pseudo wild-type fd, a revertant from a transposon-containing phage, a deletion of 64 nucleotides (pos. 5553–5618) was observed which removed

part of hairpin A and the pyrimidine-rich region between hairpins A and B (Auerswald, 1979). This region is obviously dispensable for phage multiplication. A ColE1 vector containing the left half of the IG from the end of gene IV to the *Hae*III fragment mentioned above (pos. 5489–5868) not only replicates under phage control but also packs single (plus)-strands of the vector into phage-coat protein efficiently (Sommer, 1981). The "packing origin" must therefore be localized in the left half of the IG; this indicates that hairpin A is involved in this function, as speculated earlier (Schaller, 1979).

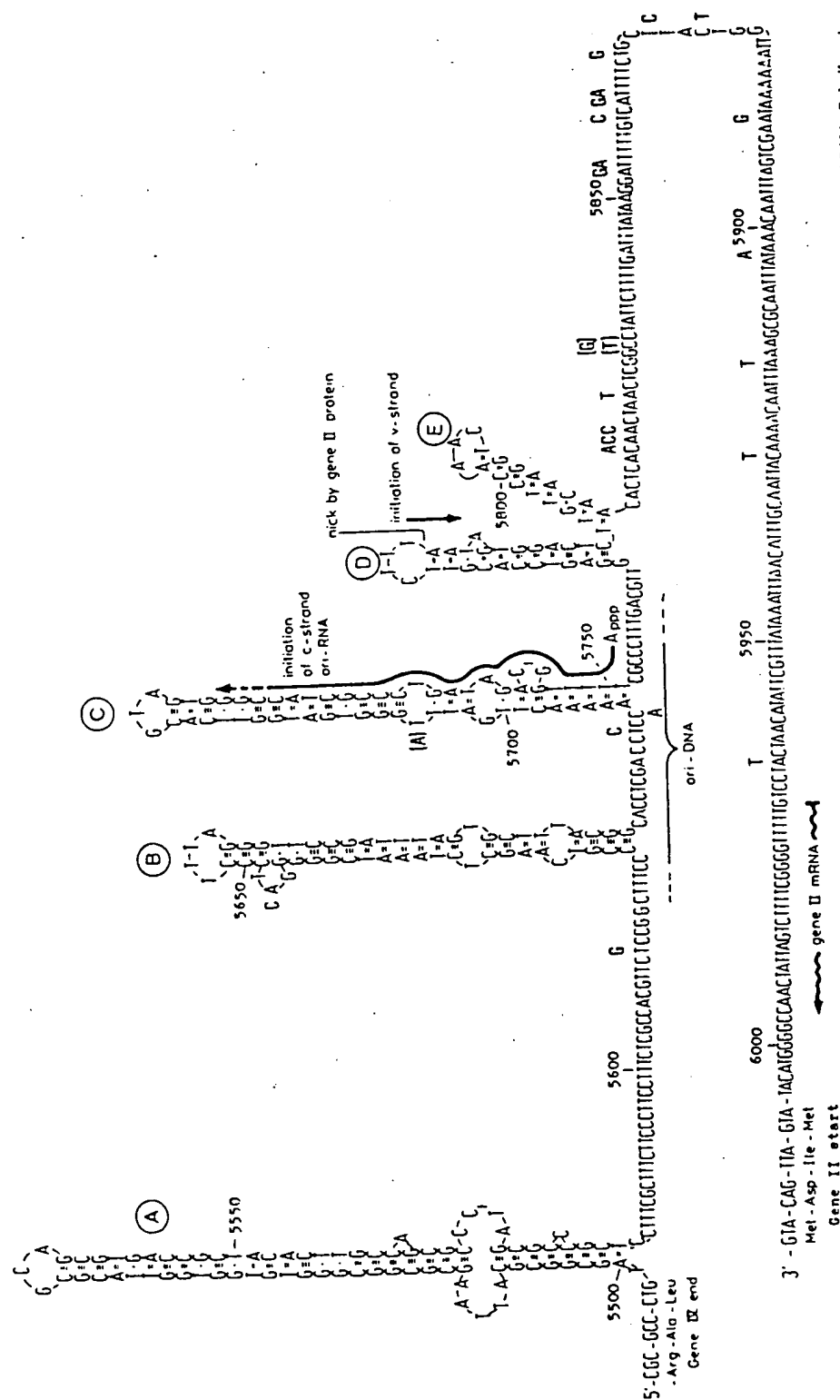
In vitro transcription starts at several promoters which are located in front of almost each gene (except of genes VII and IX) and proceeds unidirectionally to a single Rho-independent stop signal immediately after gene VIII. In this way more RNA copies are produced of the genes proximal to the central terminator than of the more distal genes. This polar effect is amplified by the fact that the strongest promoters are located in the region preceding the termination signal. The products of the genes encoded by this region are the most abundantly needed proteins of the phage. In vivo this "cascade" model of transcription could be proved only in the region between the IG and the central terminator. Some of the RNA species of this region are obviously processed in the cell (Smits et al., 1980). In the other part of the genome there exist at least two additional Rho-dependent termination signals which cease transcription behind genes VI and IV (Smits et al., 1980). In addition to the promoters known already (reviewed e.g. by Edens et al., 1976) H. Schaller (unpublished data) mapped some weaker start points of RNA synthesis. The mixture of all RNA-polymerase-binding sites protected against pDNA was isolated, 5' end-labeled and used to prime repair synthesis on fd DNA single strands. The mixture of the extended promoter regions was then cleaved with several different restriction enzymes and separated on polyacrylamide gels. By secondary cleavage with other restriction enzymes most of the resulting fragments could be positioned on the physical map. In addition, partial DNA sequences of nine from eleven binding sites isolated by this approach were determined. All these sites are listed in Fig. 4. The last nucleotide of each sequence in this figure corresponds to the first base of the (complementary) coding

TABLE III

The gene products of filamentous bacteriophages

| Gene | Start and stop codon (position) | Function and appearance | No. of amino acid residues and most frequent amino acids (%) | M_r (dalton) | Silent base exchanges fd - fl (%) | T in 3rd pos. (%) | Protein data |
|------|--|---|--|----------------|-----------------------------------|-------------------|--|
| II | ATG 6007 (90% start) ATG 6016 (10% start) | initiation of replication of viral strand DNA; nicks the plus strand of RF DNA between nucleotides pos. 5763 and 5764 | 410 Leu 11.5 Ser 11 Ala 6.8 | 46260 | 30 31 | 45 | M_r on SDS gel 46000 daltons (Konings et al., 1975); N-terminal amino acid sequence determined (Meyer et al., 1980) |
| X | ATG 496 TAA 880 | unknown; not detected <i>in vivo</i> | 111 Ser 10.8 Val 9.0 Asn 7.2 Leu 7.2 | 12680 | 10 10 | 56 | 12000 dalton protein in coupled transcription-translation systems encoded by fl DNA $HpaII$ fragment C (pos. 314 - 966) (Model and Zinder, 1974) |
| V | ATG 843 TAA 1004 | single strand specific DNA binding protein; concentration in cell ($\sim 10^5$ copies) determines the fraction of viral DNA that replicates or is packaged into capsid protein | 87 Leu 11.5 Val 9.2 Gly 8 Ser 8 | 9690 | 7 10 | 45 | amino acid sequence determined (Nakashima et al., 1974) |
| VII | ATG 1108 TCA 1207 | unknown | 33 Ile 15.2 Ala 12.1 Glu 12.1 no His | 3600 | 0 0 | 30 | |

| | | | | | | | |
|------|--|--|--|---------------------------------|----------|----|--|
| IX | ATG 1206 TGA 1302 | capsid protein few copies per phage | 32 Ser 6 Arg 2 no His | 3650 | 0 0 | 47 | amino acid compos. and M_r of C protein agrees with data derived from DNA (Simons et al., 1979) |
| VIII | ATG 1301 double stop TCATAA pos. 1520 | major capsid protein; about 1900 copies per phage (Asbeck et al., 1969) | precursor 73 mature 50 Ala 20 | prec. 7630 mat. 5240 | 1 1 | 26 | amino acid sequence of ma- ture protein determined (Asbeck et al., 1969; Na- kashima and Koningsberg, 1974) |
| III | CTG 1579 TAA 2851 | minor capsid protein; essential for adsorp- tion to the F-pilus; 5 copies per phage; (Goldsmith and Konings- berg, 1977, corrected by the real M_r) | precursor 424 mature 406 Gly 16 | prec. 44640 mat. 42609 | 14 14 | 51 | amino acid composition and N-terminal amino acid se- quence of mature protein determined; M_r from SDS gels 55000-68000 daltons (Gold- smith and Koningsberg, 1977) |
| VI | ATG 2856 TAA 3193 | capsid protein few copies per phage | 112 Leu 21.4 Ile 11.6 no Glu no His | 12350 | 4 4 | 52 | amino acid compos. and M_r of D protein agrees with data derived from DNA (Simons et al., 1979) |
| I | ATG 3197 TAA 4242 | morphogenesis | 348 Leu 10.9 Lys 9.2 Ser 8.3 Val 7.8 | 39530 | 27 28 | 50 | M_r from SDS gels 35000 - 36000 daltons (Model and Zinder, 1974; Konings et al., 1975) |
| IV | ATG 4221 TAG 5499 | morphogenesis | 426 Ser 13.6 Val 10.8 Leu 9.6 no Cys | 45780 | 61 70 | 51 | M_r from SDS gels 48000 dal- tons (Konings et al., 1975) |



Gene II start

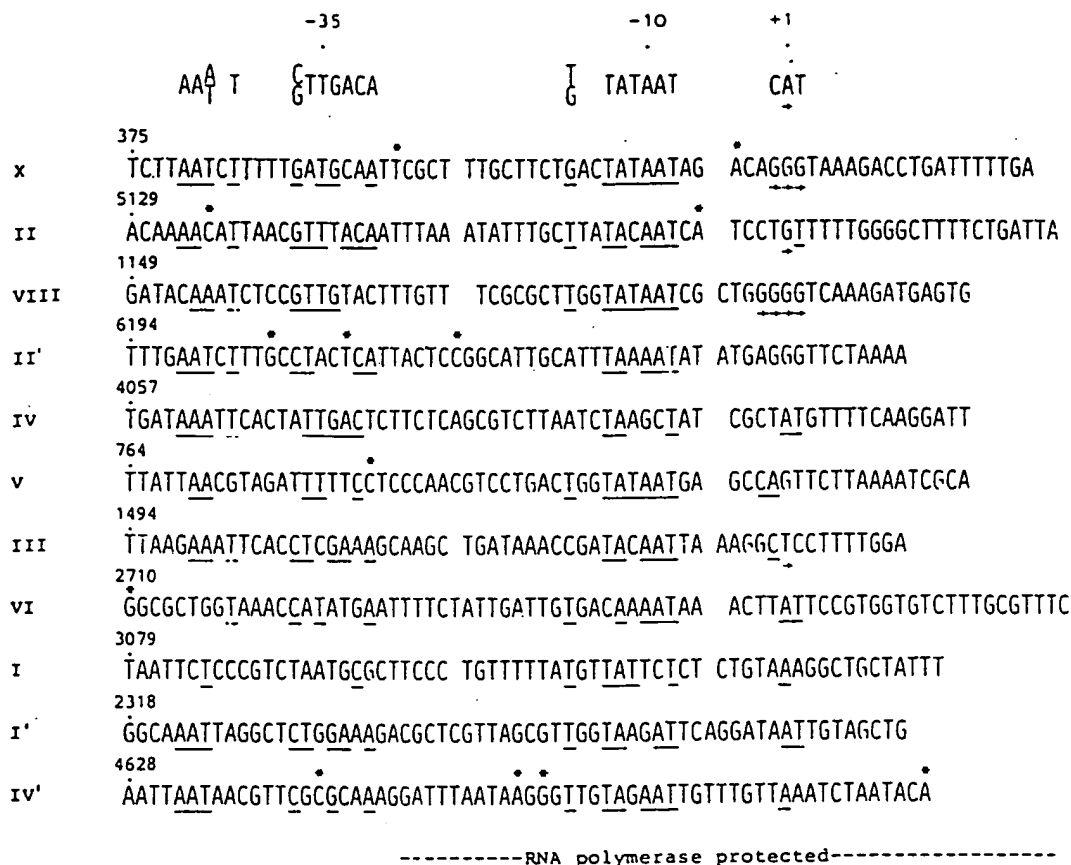


Fig. 4. Nucleotide sequences of promoter sites in fd DNA. The sequences are aligned with respect to the known initiation nucleotides (→) and the RNA polymerase recognition sites. The rightmost nucleotide in each line corresponds to the first protected base in the pDNA (minus)-strand determined by H. Schaller, as described in the text. Homologies to the consensus sequences around positions -35 and -10 (top line), as compiled by Siebenlist et al. (1980), are underlined. The upper four (strong G-start) promoters have been ordered according to their relative strength (Seeburg et al., 1977). Base exchanges to f1 and M13 DNA are marked by asterisks (see Fig. 2).

strand protected in the RNA polymerase-promoter complex.

The sequences show homologies around positions -10 and -35 to other promoter sites of *E. coli* RNA polymerase (reviews: Rosenberg and Court, 1979; Siebenlist et al., 1980). In the weaker promoters (lower part of Fig. 4) this homology is less pronounced.

In most cases promoters are integrated into the end of the preceding gene. The positions of the polymerase binding sites in front of genes II, X, V, and VII (positions 5940–5980; 400–440; 790–830, and 1170–1210, respectively) coincide with four strong

G-start promoters determined by in vitro transcription of restriction fragments with RNA polymerase (Seeburg and Schaller, 1975; Edens et al., 1976). The gene V promoter was positioned incorrectly in previous publications (Schaller et al., 1978; Van Wezenbeek et al., 1980), since a provisional fd sequence was used in the initial interpretation of the mapping of the RNA polymerase binding sites mentioned above. The end of the polymerase protected DNA was determined 105 nucleotides away from the next *Hin*fl cleavage site, which is at position 723 and not at position 741. The thus derived polymerase-binding site shows a perfect Pribnow

TABLE IV

Amber mutants of fd, f1 and M13

| Gene | Phage | Name of mutant | Position | Base exchange |
|------|-------|-----------------------------------|----------|---------------|
| II | f1 | R124 | 6349 | G → T |
| V | fd | fd122 ^a | 906 | C → T |
| | M13 | 5H1 ^a 5H3 ^a | 999 | C → T |
| | | 5H27 ^a | | |
| | f1 | R13 | 999 | C → T |
| VII | M13 | 7H2 ^b | 1114 | C → T |
| | M13 | 7H3 ^b | 1141 | C → T |
| VIII | M13 | 8H1 ^c | 1373 | G → T |
| III | M13 | 3H1 ^a 3H4 ^a | 2017 | C → T |
| | M13 | 3H5 ^a | 2473 | C → T |
| VI | f1 | R5 R7 | 3066 | C → T |
| | M13 | 6H1 ^a 6H2 ^a | 3066 | C → T |
| | | 6H3 ^a 6H6 ^a | | |
| I | M13 | 1H7 ^a | 3263 | C → T |
| IV | f1 | R143 | 5265 | C → T |

^a Van Wezenbeek et al. (1980); ^b Hulsebos and Schoenmakers (1978); ^c Boeke and Model (1979).

hexamer. A corrected assignment of this promoter in the fd DNA sequence is presented (Siebenlist et al., 1980), but also has to be shifted in the -35 region by one nucleotide, since in M13, as well as in f1, there is a T at position 783 instead of a C. An exchange of the corresponding C in position -32 for a T causes a down mutation in the λp_L promoter (sex1 mutant; Kleid et al., 1976). The alignment to the -35 region proposed here results in an equivalent homology pattern as that shown by Siebenlist et al. (1980), with the base exchange at a point of nonhomology. In the II' promoter, which is a strong RNA start site in vitro, position -33 is converted by a T → A base exchange in f1 and M13 (pos. 6213) into the "ideal" form. There is no mutant known in this position in other promoters, but the potentially different frequency of RNA initiation at this site in fd on the one hand and in f1 and M13 on the other hand is not yet measured.

There are several other changes within the promoters of the filamentous phages, indicated in Fig. 3. All of them concern positions outside the polymerase recognition sites except for a C → G exchange in the -35 region of promoter IV'. It is

TABLE V

Amino acid exchanges in proteins between fd, f1 and M13

| Gene | Amino acid pos. | Amino acid | | | Pos. of exchange in DNA |
|------|-----------------|------------|------------------|-----|-------------------------|
| | | fd | f1 | M13 | |
| II | 249 | Glu | Glu | Lys | 343 |
| | 274 | Arg | Ser | Ser | 420 |
| VIII | 35 | Asp | Asp | Asn | 1403 |
| III | 374 | Pro | Leu ^a | Pro | 2699 ^a |
| | 375 | Tyr | Phe | Phe | 2702 |
| | 378 | Gly | Arg | Ser | 2710 |
| I | 142 | His | Asn | His | 3620 |
| | 164 | Val | Ile | Ile | 3686 |
| | 326 | Ile | Leu | Leu | 4172 |
| IV | 30 | Pro | Ser | Pro | 4308 |
| | 42 | Thr | Thr | Ser | 4344 |
| | 70 | Asn | Asp | Asn | 4428 |
| | 98 | Ser | Asn | Asn | 4513 |
| | 110 | Ile | Asn | Asn | 4549 |
| | 166 | Val | Val | Ile | 4716 |

^a This exchange was observed only in f1 amber mutant R5.

questionable whether this promoter and the I' promoter have any function in vivo since these sites were only detected as polymerase binding sites in vitro and not by a transcription product.

The position of the gene VIII promoter was also established by sequence analysis of the gene transcript, which starts with pppG₄ at position 1196 (Takanami et al., 1976). Gene IX is encoded by the same mRNA. The ATG start codon lies 10 nucleotides downstream from the beginning of the mRNA. Transcripts probably do not start exclusively at one single position: like the "wobbling" start of gene X mRNA (Nüsslein and Schaller, 1975), a percentage may initiate either one nucleotide before or one after, giving rise to varying numbers of G residues at the 5' end. Only the longest RNA chains starting with G5 may offer an efficient ribosome-binding site, which perhaps accounts for the low expression of gene IX.

Three RNA polymerase binding sites in front of genes VI, I, and IV (pos. 2740-2780, 3100-3140, and 4080-4120, respectively) confirm the positions of three A-start promoters, also determined by in vitro transcription of restriction fragments (Edens et al., 1976). A further binding site (pos.

1510–1550) overlaps partially with the central termination signal for transcription and defines the position of the gene III promoter. The mRNA probably starts at position 1544 with pppU (M. Takanami, personal communication; Edens et al., 1978).

Sequences preceding the start codons of the genes listed in Fig. 5 show varying degrees of complementarity to the 3' end of the 16s rRNA (Shine and Dalgarno, 1974). Three ribosome-binding sites of

phage f1 were isolated as early as 1974 from ribosome-RNA complexes, and the nucleotide sequence was analysed (Pieczenik et al., 1974). In establishing the DNA sequence it appeared that these sites belonged to the genes V, VI, and VIII.

The start codons for all genes are ATG except for gene III, which starts with GTG, possibly contributing to the low expression rate of this gene. In genes that are efficiently expressed (e.g. genes II, V, and VIII) an A follows the ATG codon, which is in

| | |
|-----------|---|
| 16s rRNA | 3' OH AUUCCUCCACUAG-- |
| Gene II | 5991 ATCAACCGGGGTACAT ATG ATT GAC ATG CTA |
| Gene II' | 6000 GGTACATATGATTGAC ATG CTA GTT TTA CGA |
| Gene X | 480 ATTTCGCGGGGGATTCA ATG AAT ATT TAT GAC |
| Gene V | 827 CATAGGCTAATTCAAA ATG ATT AAA GTT GAA |
| Gene VII | 1097 GTTCCGGCTAATCTAC ATG GAG CAG GTC GCG |
| Gene IX | 1190 TCGCTGGGGGTCAAAG ATG AGT GTT TTA GTG |
| Gene VIII | 1285 TAATGGAACTTCCTC ATG AAA AAG TCT TTA |
| Gene III | 1563 TTTGGAGATTTTCAAC GTG AAA AAA TTA TTA |
| Gene VI | 2840 ATAAGGAGTCTTAAATC ATG CCA GTT CTT TTG |
| Gene I | 3181 ATTGGGATATAATAT ATG GCT GTT TAT TTT |
| Gene IV | 4205 AAAAAAGCTAATTCAA ATG AAA TTG TTA AAT |

Fig. 5. Nucleotide sequences of ribosome binding sites in fd DNA. Nucleotides complementary to the 3'-terminus of 16s rRNA (Shine and Dalgarno, 1974) are underlined. Palindrome structures near the start codon are indicated by arrows, and stop signals preceding the start codons are boxed.

agreement with the hypothesis that the fourth base in the f-Met-tRNA anticodon is involved in the formation of the translation-initiation complex (Taniguchi and Weissmann, 1978).

In all these ribosome binding sites a palindrome can be observed more or less evidently (indicated by arrows in Fig. 5), which allows part of the sequence upstream from the start codon to base-pair with the sequence downstream, thus exposing the ATG triplet on top of a small hairpin structure. Such structures were first considered as possible translation recognition signals in other systems (Steitz and Jakes, 1975), but this idea was later rejected (Steitz, 1979). Similar structures are also recognized at other ribosome-binding sites (coat and A proteins of $\phi 2$, MS2, Q β , genes C and F of ϕ X174, genes *lacI*, *galE*, *galT* of *E. coli*) as listed in Steitz (1979).

Stop codons immediately precede or overlap with translational start signals due, primarily, to the close packing of genes in the filamentous phage genome (see above). However, this arrangement may also provide a helper function for translation-promoted re-initiation of translation: The ribosomes stop in a position that allows Shine-Dalgarno base pairing to occur anew.

The completed ϕ I DNA sequence shows this phage to be closely related to the two other filamentous bacteriophage ϕ d and M13. The small gene products are almost all identical to their various counterparts, whereas the amino acid sequences of the larger proteins diverge from one another by as much as 2%. Regulatory elements also vary only slightly in their essential parts. More variable regions lie in the IG between highly conserved segments, the latter probably representing structurally functional domains. Such variable regions can, in part, be deleted or replaced by heterologous DNA, which allowed the filamentous phage to be used as efficient cloning vehicles (Messing et al., 1977; Herrmann et al., 1980).

ACKNOWLEDGEMENTS

We thank Dr. H. Schaller for discussion, Dr. M. Takanami for unpublished data, R. Sommer for help with the computer processing of sequence data and P. Shaw for help with the manuscript. This work was

supported by a grant from the Deutsche Forschungsgemeinschaft (Scha 134/8).

REFERENCES

- Asbeck, F., Beyreuther, K., Köhler, H., von Wettstein, G. and Braunitzer, G.: Virusproteine, IV. Die Konstitution des Hüllproteins des Phagen ϕ d. Hoppe Seylers Z. Physiol. Chem. 350 (1969) 1047-1066.
- Auerswald, E.A.: Struktur des Transposons Tn5 und Analyse seiner Integration und Exzision im Bakteriophagen ϕ d. Ph. D. Thesis, University of Heidelberg, 1979.
- Beck, E., Sommer, R., Auerswald, E.A., Kurz, Ch., Zink, B., Osterburg, G., Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M.: Nucleotide sequence of bacteriophage ϕ d DNA. Nucl. Acids Res. 5 (1978) 4495-4504.
- Blakesley, R.W. and Wells, R.D.: Single-stranded DNA from ϕ X174 and M13 is cleaved by certain restriction endonucleases. Nature 257 (1975) 421-422.
- Boeke, J.D. and Model, P.: Molecular basis of the am8HI lesion in bacteriophage M13. Virology 96 (1979) 299-301.
- Cleary, J.M. and Ray, D.S.: Replication of the plasmid pBR322 under the control of a cloned replication origin from the single-stranded DNA phage M13. Proc. Natl. Acad. Sci. USA 77 (1980) 4638-4642.
- Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G.: Transcription of bacteriophage M13 DNA: Existence of promoters directly preceding genes III, VI and I. J. Virol. 28 (1978) 825-842.
- Edens, L., Van Wezenbeek, P., Konings, R.N.H. and Schoenmakers, J.G.G.: Localization of promoter regions on the genome of bacteriophage M13. Eur. J. Biochem. 70 (1976) 577-587.
- Fuchs, C., Rosenfold, E.C., Honigman, A. and Szybalski, W.: Identification of palindromic sequences by restriction endonucleases, as based on the tabularized sequencing data for seven viral and plasmid DNAs. Gene 10 (1980) 357-370.
- Geider, K., Beck, E. and Schaller, H.: An RNA transcribed from DNA of the origin of phage ϕ d single-strand to replicative form conversion. Proc. Natl. Acad. Sci. USA 75 (1978) 645-649.
- Godson, G.N., Barrell, B.G., Staden, R. and Fiddes, J.C.: Nucleotide sequence of bacteriophage G4 DNA. Nature 276 (1978) 236-247.
- Goldsmith, M.E. and Koningsberg, W.H.: Adsorption protein of the bacteriophage ϕ d: Isolation, molecular properties and location in the virus. Biochemistry 16 (1977) 2686-2694.
- Gray, C.P., Sommer, R., Polke, C., Beck, E. and Schaller, H.: Structure of the origin of DNA replication of bacteriophage ϕ d. Proc. Natl. Acad. Sci. USA 75 (1978) 50-53.

- Herrmann, R., Neugebauer, K., Pirkli, E., Zentgraf, H. and Schaller, H.: Conversion of bacteriophage fd into an efficient single-stranded DNA vector system. *Mol. Gen. Genet.* 177 (1980) 231-242.
- Herrmann, R., Neugebauer, K., Zentgraf, H. and Schaller, H.: Transposition of a DNA sequence determining kanamycin resistance into the single-stranded genome of bacteriophage fd. *Mol. Gen. Genet.* 159 (1978) 171-178.
- Hulsebos, T. and Schoenmakers, J.G.G.: Nucleotide sequence of gene VII and of a hypothetical gene (IX) in bacteriophage M13. *Nucl. Acids Res.* 5 (1978) 4677-4698.
- Johnson, R.A. and Walseth, T.F.: The enzymatic preparation of [α - 32 P]ATP, [α - 32 P]GTP, [32 P]cAMP, [32 P]cGMP and their use in the assay of adenylate and guanylate cyclases and cyclic nucleotide phosphodiesterases. *Adv. Cycl. Nucleotide Res.* 10 (1979) 135-167.
- Kleid, D., Humayun, Z., Jeffrey, A. and Ptashne, M.: Novel properties of a restriction endonuclease isolated from *Haemophilus parahaemolyticus*. *Proc. Natl. Acad. Sci. USA* 73 (1976) 293-297.
- Konings, R.N.H., Hulsebos, T. and Van den Hondel, C.A.: Identification and characterisation of the in vitro synthesized gene products of bacteriophage M13. *J. Virol.* 15 (1975) 570-584.
- Ling, V.: Fractionation and sequences of the large pyrimidine oligonucleotides from bacteriophage fd DNA. *J. Mol. Biol.* 64 (1972) 87-102.
- Lyons, L.B. and Zinder, N.D.: The genetic map of the filamentous bacteriophage f1. *Virology* 49 (1972) 45-60.
- Marvin, D.A. and Hohn, B.: Filamentous bacterial viruses. *Bacteriol. Rev.* 33 (1969) 172-209.
- Maxam, A.M. and Gilbert, W.: A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* 74 (1977) 560-564.
- Maxam, A.M. and Gilbert, W.: Sequencing end-labeled DNA with base-specific chemical cleavages, in Grossman, L. and Moldave, K. (Eds.), *Methods in Enzymology*, Vol. 65. Academic Press, New York, 1980, pp. 499-560.
- Messing, J., Gronenborn, B., Müller-Hill, B. and Hofschneider, P.H.: Filamentous coliphage M13 as a cloning vehicle. Insertion of a *Hind*III fragment of the *lac* regulatory region in M13 replicative form in vitro. *Proc. Natl. Acad. Sci. USA* 74 (1977) 3642-3646.
- Meyer, T.F., Beyreuther, K. and Geider, K.: Recognition of two initiation codons for the synthesis of phage fd gene II protein. *Mol. Gen. Genet.* 180 (1980) 489-494.
- Meyer, T.F., Geider, K., Kurz, Ch. and Schaller, H.: Cleavage site of bacteriophage fd gene II protein in the origin of viral strand replication. *Nature* 278 (1979) 365-366.
- Model, P. and Zinder, N.D.: In vitro synthesis of bacteriophage f1 proteins. *J. Mol. Biol.* 83 (1974) 231-251.
- Nakashima, Y., Dunker, A.K., Marvin, D.A. and Koningsberg, W.: The amino acid sequence of a DNA binding protein, the gene 5 product of fd filamentous bacteriophage. *FEBS Lett.* 43 (1974) 125.
- Nakashima, Y. and Koningsberg, W.: Reinvestigation of a region of the fd bacteriophage coat protein sequence. *J. Mol. Biol.* 88 (1974) 598-600.
- Nüsslein, C. and Schaller, H.: Stabilisation of promoter complexes with a single ribonucleoside triphosphate. *Eur. J. Biochem.* 56 (1975) 563-569.
- Oertel, W. and Schaller, H.: A new approach to the sequence analysis of DNA. *FEBS Lett.* 27 (1972) 316-320.
- Ohmori, H., Tomizawa, J. and Maxam, A.M.: Detection of 5-methylcytosine in DNA sequences. *Nucl. Acids Res.* 5 (1978) 1479-1485.
- Osterburg, G. and Sommer, R.: Computer support of DNA sequence analysis. *Comp. Prog. Biomed.* 13 (1981) 101-109.
- Piecznik, G., Model, P. and Robertson, H.D.: Sequence and symmetry in ribosome-binding sites of bacteriophage f1 RNA. *J. Mol. Biol.* 90 (1974) 191-214.
- Ravetch, J.V., Horiuchi, K. and Zinder, N.D.: Nucleotide sequences near the origin of replication of bacteriophage f1. *Proc. Natl. Acad. Sci. USA* 74 (1977) 4219-4223.
- Ravetch, J.V., Horiuchi, K. and Zinder, N.D.: DNA sequence analysis of the defective interfering particle of bacteriophage f1. *J. Mol. Biol.* 128 (1979) 305-318.
- Ray, D.S.: Replication of filamentous bacteriophages, in Fraenkel-Conrat, H. and Wagner, R. (Eds.), *Comprehensive Virology*, Vol. 7. Plenum, New York, 1977, pp. 105-178.
- Roberts, R.J., Myers, P.A., Morrison, A. and Murray, K.: A specific endonuclease from *Haemophilus haemolyticus*. *J. Mol. Biol.* 103 (1976) 199-208.
- Rosenberg, M. and Court, D.: Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* 13 (1979) 319-353.
- Sanger, F., Air, G., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchinson, C.A., Slocombe, P.M. and Smith, M.: Nucleotide sequence of bacteriophage ϕ X174 DNA. *Nature* 265 (1977) 687-695.
- Sanger, F., Coulson, A.R.: A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 94 (1975) 441-448.
- Sanger, F., Donelson, J.E., Coulson, A., Kössel, H., Fischer, D.: Use of DNA Polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage f1. *Proc. Natl. Acad. Sci. USA* 70 (1973) 1209-1213.
- Sanger, F., Donelson, J.E., Coulson, A., Kössel, H., Fischer, D.: Determination of a nucleotide sequence in bacteriophage f1 DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 90 (1974) 315-333.
- Schaller, H.: The intergenic region and the origins for filamentous phage DNA replication. *Cold Spring Harbor Symp. Quant. Biol.* 43 (1979) 401-408.
- Schaller, H., Beck, E. and Takanami, M.: Sequence and regulatory signals of the filamentous phage genome, in Denhardt, D.T., Dressler, D., and Ray, D.S. (Eds.), *The Single-Stranded DNA Phages*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 1978, pp. 139-163.
- Schaller, H., Gray, C.P. and Herrmann, R.: Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd. *Proc. Natl. Acad. Sci. USA* 72 (1975) 737-741.
- Schaller, H., Uhlmann, A. and Geider, K.: A DNA fragment

- from the origin of single-strand to double-strand replication of bacteriophage fd. *Proc. Natl. Acad. Sci. USA* 73 (1976) 49-53.
- Scherer, G.E.F., Walkinshaw, M.D. and Arnot, S.: A computer aided oligonucleotide analysis provides a model for RNA polymerase-promoter recognition. *Nucl. Acids Res.* 5 (1978) 3759-3773.
- Seeburg, P.H., Nüsslein, C. and Schaller, H.: Interaction of RNA polymerase with promoters from bacteriophage fd. *Eur. J. Biochem.* 74 (1977) 107-113.
- Seeburg, P.H. and Schaller, H.: Mapping and characterization of promoters in phage fd, f1 and M13. *J. Mol. Biol.* 92 (1975) 261-277.
- Shine, J. and Dalgarno, L.: The 3'-terminal sequence of *Escherichia coli* 16s ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* 71 (1974) 1342-1346.
- Siebenlist, U., Simpson, R.B. and Gilbert, W.: *E. coli* RNA polymerase interacts homologously with two different promoters. *Cell* 20 (1980) 269-281.
- Simons, G.F.M., Konings, R.N.H. and Schoenmakers, J.G.G.: Identification of two new capsid proteins in bacteriophage M13. *FEBS Lett.* 106 (1979) 8-12.
- Smiths, M.A., Schoenmakers, J.G.G. and Konings, R.N.H.: Expression of bacteriophage M13 DNA in vivo, IV. Isolation, identification and characterization of phage-specific mRNA species. *Eur. J. Biochem.* 112 (1980) 309-321.
- Sommer, R.: Klonierung und Analyse der für die Replikation und Phagenmorphogenese essentiellen DNA Strukturen des Bakteriophagen fd. Ph. D. Thesis, University of Heidelberg, 1981.
- Steitz, J.A.: Genetic signals and nucleotide sequences in messenger RNA, in Goldberg, R.F. (Ed.), *Biological regulation and development, I. Gene Expression*. Plenum, New York, 1979, pp. 349-399.
- Steitz, J.A. and Jakes, K.: How ribosomes select initiator regions in mRNA: Base pair formation between the 3'-terminus of 16s rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 72 (1975) 4734-4738.
- Suggs, S.V. and Ray, D.S.: Nucleotide sequence of the origin for bacteriophage M13 DNA replication. *Cold Spring Harbor Symp. Quant. Biology* 43 (1978) 379-388.
- Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M.: Studies on bacteriophage fd DNA, III. Nucleotide sequence preceding the RNA start site on a promoter-containing fragment. *Nucl. Acids Res.* 2 (1975) 2091-2100.
- Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M.: Studies on bacteriophage fd DNA, IV. The sequence of messenger RNA for the major coat protein gene. *J. Mol. Biol.* 111 (1977) 487-507.
- Takanami, M., Sugimoto, H., Sugisaki, H. and Okamoto, T.: Sequence of the promoter for the coat protein gene of the bacteriophage fd. *Nature* 260 (1976) 297-302.
- Taniguchi, T. and Weissmann, C.: Site-directed mutations in the initiator region of the bacteriophage Q β coat cistron and their effect on ribosome binding. *J. Mol. Biol.* 118 (1978) 533-565.
- Van Wezenbeek, P.M.G.F., Hulsebos, T.J.M. and Schoenmakers, J.G.G.: Nucleotide sequence of the filamentous bacteriophage M13 genome: Comparison with phage fd. *Gene* 11 (1980) 129-148.
- Vovis, G.F., Horiuchi, K. and Zinder, N.D.: Endonuclease R. *EcoRII* restriction of bacteriophage f1 DNA in vitro: Ordering of genes V and VII, location of an RNA promoter for gene VIII. *J. Virol.* 16 (1975) 674-684.

Communicated by W. Fiers.